

Zápočtový test – 19.12.2007

V dotazníku byly položeny dvě otázky

11. Kouříš?

ne, nikdy 101□ pravidelně 103□
příležitostně 102□ bývalý kuřák 104□

12. Piješ alkohol?

abstinent 101□ 1x týdně 103□
příležitostně 102□ vícekrát týdně 104□

Proveďte tyto výpočty, u každého příkladu uveďte krátké slovní hodnocení

Zvolte vhodnou hladinu významnosti α . Jednou větou napište, co to znamená.

Většina z vás zvolila hladinu významnosti $\alpha = 0,1$. Vzhledem k počtu respondentů (téměř 2000) mohla být zvolena klidně i 8 %-ní ($\alpha = 0,08$). Úvaha vedoucí k nastavení α může být následující. Mám 2000 dotazníků. Jaké procento chyby jsem ochoten tolerovat? Z těch 2000 lidí klidně 150 respondentů může odpovědět jinak, než jsou mé hypotézy a tento počet je pro mě stále vyhovující (protože 1850 odpoví v souladu s hypotézou). Alfou vypočítám: $150 / 2000 * 100 = 0,075$. Z toho vyplývá $\alpha = 0,08$.

Každopádně $\alpha = 0,1$ nebo $\alpha = 0,05$ (nebo i jiná) je pouze na řešiteli a obě hodnoty jsou správné.

a) Základní informace o souboru – aritmetický průměr, směrodatnou odchylku, medián + DISKUSI k výsledku

Zde byly většinou správně vypočtena čísla, ale málo lidí vysvětlilo, co znamená, že na otázku „kouříš“ odpověděli lidé průměrně 101,6. V zadání je napsáno, že pod kódem 101 je odpověď „ne, nikdy“, pod 102 „příležitostně“ atd. Tedy průměrná odpověď 101,6 znamená odpověď mezi „ne, nikdy“ a „příležitostně“. Takže v souhrnu respondenti byli spíše nekuřáci.

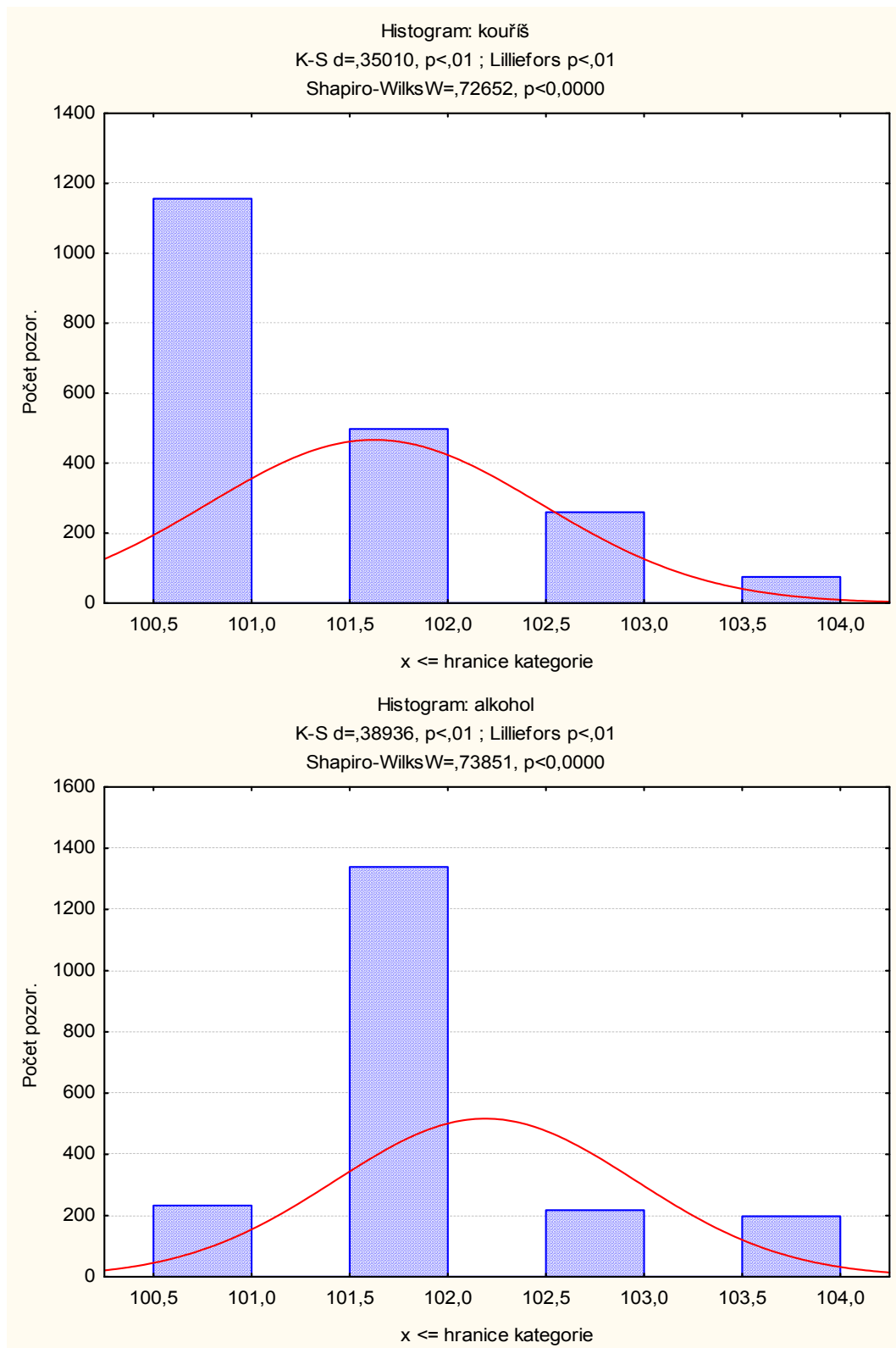
S alkoholem udávají výsledky, že respondenti pijí alkohol příležitostně.

Proměnná	Popisné statistiky (np015-data)		
	Průměr	Medián	Sm. odch.
kouříš	101,6249	101,0000	0,849962
alkohol	102,1916	102,0000	0,767345

b) Testy normality

Pocházejí data z normálního rozdělení? Jestli ano/ne, uveďte na základě čeho jste se rozhodli, vložte graf či tabulku, ze které to bude patrné.

Na základě Kolmogorova-Smirnovova testu nebo Shapiro-Wilksova testu zamítáme hypotézu o normalitě dat. ($p = 0$, což je menší než α). Viz histogramy a údaje v záhlaví grafů



c) Zjistěte míru závislosti mezi dvěma proměnnými (korelace), proveďte diskusi. Jde o jakou úměru (přímou/nepřímou)?

Korelační koeficient mezi oběma proměnnými je 0,35. Hodnota korelačního koeficientu je významná na hladině významnosti $\alpha = 0,1$ i $\alpha = 0,05$ ($p=0$). Hodnota je kladná, jedná se tedy o přímou úměrnost. Velikost koeficientu není moc vysoká, čili můžeme konstatovat, že existuje vztah mezi kouřením a pitím alkoholu, ale nelze tento vztah považovat za nějak průkazný.

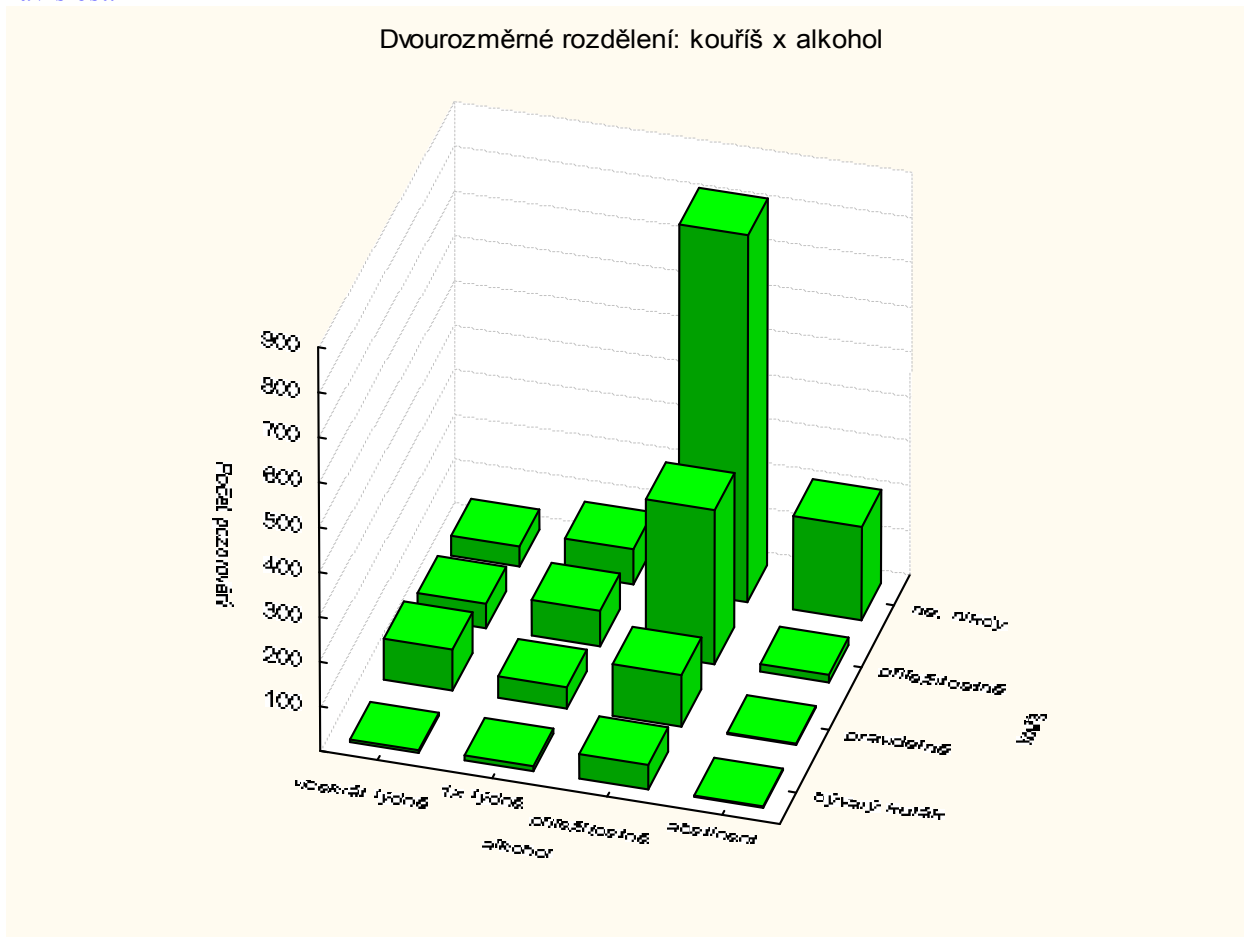
Korelace (np015-data) Označ. korelace jsou významné na hlad. $p < ,05000$ N=1988 (Celé případy vynechány u ChD)		
Proměnná	kouříš	alkohol
kouříš	1,0000	,3498
	p= ---	p=0,00
alkohol	,3498	1,0000
	p=0,00	p= ---

d) Zjistěte míru nezávislosti mezi dvěma proměnnými (test nezávislosti chí-kvadrát), včetně grafu, proveďte diskusi

Nejprve se podívejme na kontingenční tabulku = tabulka četností.

Kontingenční tabulka (np015-data) Četnost označených buněk > 10 (Marginální součty nejsou označeny)					
kouříš	alkohol abstinent	alkohol příležitostně	alkohol 1x týdně	alkohol vícekrát týdně	Řádk. součty
ne, nikdy	210	822	79	45	1156
příležitostně	17	346	80	55	498
pravidelně	3	116	48	92	259
bývalý kuřák	3	55	11	6	75
Vš.skup.	233	1339	218	198	1988

Testováním hypotézy o nezávislosti pomocí Pearsonova chí-2 koeficientu zamítáme hypotézu o nezávislosti. Čili zde existuje závislost. Můžeme se pokusit ji vysvětlit např. tak, že typické zastoupení respondentů jsou nekuřáci a příležitostní pijaní alkoholu. U ostatních kategorií sledovaných znaků neexistuje výrazná závislost.



Statist.	Statist. : kouříš(4) x alkohol(4) (np015-data)		
	Chí-kvadr.	sv	p
Pearsonův chí-kv.	379,2711	df=9	p=0,0000
M-V chí-kvadr.	348,4120	df=9	p=0,0000

e) Porovnejte, která z proměnných dosahuje vyššího skóre

Použití t-testu a to neparametrického Wilcoxonova testu. Proč neparametrického? Protože jsme ověřili, že ani jedna proměnná nepochází z normálního rozdělení. Proč Wilcoxonův test? Protože se jedná o párová pozorování (jeden člověk odpověděl na dvě otázky)

Dvojice proměnných	Wilcoxonův párový test (np015-data)			
	Počet platných	T	Z	Úroveň p
kouříš & alkohol	1988	157672,5	21,48215	0,00

Na základě výsledků Wilcoxonova testu zamítáme hypotézu o rovnosti středních hodnot odpovědí na sledované dvě otázky a tvrdíme, že mezi nimi existuje statisticky významný rozdíl. A to ve prospěch odpovědi „alkohol“. Viz ad a)

Seznam nulových hypotéz:

- Hypotéza o normalitě souboru
- Hypotéza o nulovosti korelačního koeficientu
- Hypotéza o nezávislosti dvou proměnných (chí-kvadrát)
- Hypotéza o rovnosti středních hodnot dvou výběrů (t-test)

Pro úspěšné ukončení je třeba z 5 příkladů vypočítat 3.